U. S. DEPARTMENT OF COMMERCE
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
NATIONAL WEATHER SERVICE
NATIONAL METEOROLOGICAL CENTER

OFFICE NOTE 85

(REVISED)

The Standard Label for NMC Data Sets

Automation Division Staff

December 1990

PURPOSE

This Office Note describes the format to be used in constructing the internal Data Set Label that should appear in the first record of any and all data sets (files) generated during production runs at NMC. The label should be present in the data set whether the data set resides on magnetic or optical disk, cartridge, tape, or any other media. It contains an assortment of information needed to identify the contents of the data set, its format when originally constructed, and when and where it was created.

The Data Set Label is used operationally to identify routine inventories, help track problems, and allow for the generation of archive files that can be used to reconstruct data sets for later study.

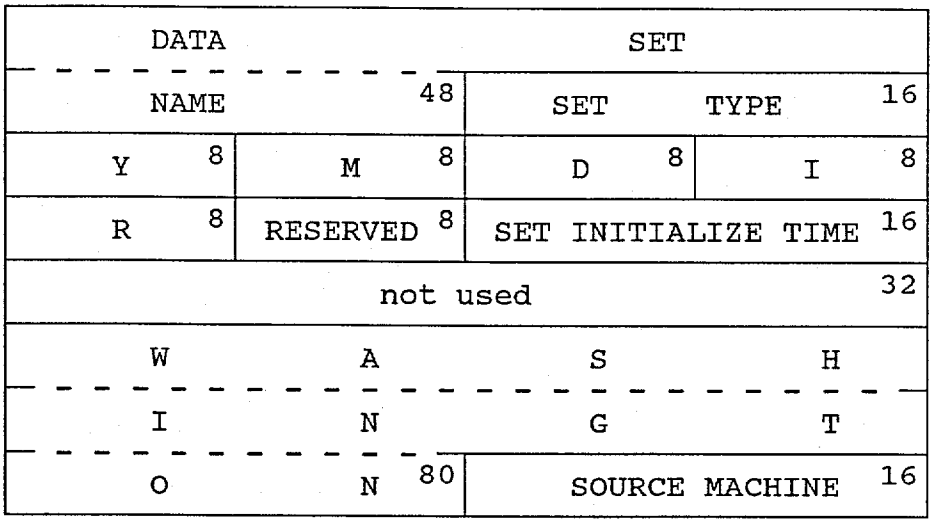
There are two broad categories of data sets produced routinely at NMC: one containing information represented in EBCDIC (Extended Binary Coded Decimal Interchange Code), often referred to as "Coded" or "Character" data, the other containing purely binary data. The binary data may be in fixed point, floating point, or any number of special packing formats. The labels for these two types of data sets differ in format although they contain essentially the same information.

BINARY SETS

The first 32 bytes of the first record on the data set (disregarding any system generated labels) contain the data set information. If the data set is a random access data set then that first record will also contain the identification of the fields (records) in the data set and pointers used by the appropriate random access I/O programs. There may be other specialized uses of the record in other applications but we concern ourselves here with the first 32 bytes only.

Here is what the label looks like:

(The little numbers count the bits in each subsection of the label.)



Where ...

DATA SET NAME: Up to six EBCDIC characters, left justified with blank fill. Conventionally, for production data sets, this name should be the third qualifier of the JCL Data Set Name. See the NMC Handbook, Section 3.6.2, for the naming conventions for those data sets.

SET TYPE: two EBCDIC characters indicating how the data set was originally created:

B1: binary random access data sets identified and packed according to NMC Office Note 84 and written with the (now unused) FORTXDAM I/O method (W3FK01 - W3FK11).

B2: binary sequential data sets.

B3: binary random access data sets identified and packed according to NMC Office Note 84 and written with the VSAM I/O method (W3FK40 - W3FK51).

B4: binary random access data sets identified and packed according to WMO Code FM 92 GRIB and written with the VSAM I/O method (W3FK70 - W3FK81).

Y: year of century (binary)

M: month (binary)

D: day (binary)

I: analysis time or initial time of forecast (binary) in units of UTC ("Greenwich" or "Z") hours - 00 through 23.

R: run number - see NMC Office Note 84.

RESERVED: Space that may be used by the I/O system to indicate splits of very long identification tables.

SET INITIALIZE TIME: The time (binary) when the data set was initialized. The units are hundredths of hours; e.g. if the data set was initialized at 15 hours 45 minutes UTC, equivalent to 15.75 UTC, the time would be entered as 1575.

WASHINGTON: The ten EBCDIC characters "WASHINGTON" (what else?).

SOURCE MACHINE: Two EBCDIC characters indicating the particular computer that created the information in the data set:

00: Front End Machine (NAS, IBM, whatever)

05: CYBER 205

CR: CRAY YMP

"Not used" fields should be set to binary zeros at the time the data set in initialized.

The format of fixed or floating point data on any device is application dependent. Often, copies of CRAY or CYBER generated files may be found on the front-end machines still in their native formats; on other occasions the format may have been changed to fit the resident machine. The 32 byte binary file label should not be taken to indicate "format" at this level of detail.

Special notes related to archives or other copies of data sets: If a random access data set (type B1, B3, B4, etc.) is copied to a sequential data set (a tape, for example) the SET TYPE parameter is NOT changed, even though the copy is inherently sequential. This is so that archive recovery programs can recreate the data set in its original form when called upon to do so. Generally, binary data formats from the CYBER and CRAY are converted to IBM format for number representation before transferring them to NMC's archive media.

The first 32 bytes of the first record on the data set (disregarding any system generated labels) contain the data set information. In coded data sets, those 32 bytes not infrequently constitute the entire record. Provision has been made for random access coded data sets even though none are currently in use nor is any such anticipated.

Here is what the label looks like:

(The little numbers count the bits in each subsection of the label.)

DATA		SET	
NAME	48	SET TYPE	16
Y	16	M	16
D	16	I	16
SET INITIALIZE TIME			32
W	A	S	H
I	N	G	T
O	N 80	SOURCE MACHINE	16

Where...

DATA SET NAME: Up to six EBCDIC characters, left justified with blank fill. Conventionally, for production data sets, this name should be the third qualifier of the JCL Data Set Name. See the NMC Handbook, Section 3.6.2, for the naming conventions for those data sets.

SET TYPE: two EBCDIC characters indicating how the data set was originally created:

C1: coded random access data sets written with the (now unused) FORTXDAM I/O method.

C2: coded sequential data sets written with FORTRAN I/O methods.

C3: coded sequential data sets written with the VSAM I/O method.

Y: year of century (2 EBCDIC characters)

M: month (2 EBCDIC characters)

D: day (2 EBCDIC characters)

I: analysis time or initial time of forecast (2 EBCDIC characters) in units of UTC ("Greenwich" or "Z") hours.

SET INITIALIZE TIME: The time (4 EBCDIC characters) when the data set was initialized. The units are hundredths of UTC hours; e.g. if the data set was initialized at 1545 UTC, which is equivalent to 15.75 UTC, the time would be entered as "1575".

WASHINGTON: The ten EBCDIC characters "WASHINGTON" (what else?).

SOURCE MACHINE: Two EBCDIC characters indicating the particular computer that created the information in the data set:

00: Front End Machine (NAS, IBM, whatever)

05: CYBER 205

CR: CRAY YMP